https://doi.org/10.2352/ISSN.2470-1173.2020.10.IPAS-312

GENERATIVE ADVERSARIAL NETWORKS: A SHORT REVIEW

Habib Ullah[1] *Sultan Daud Khan*[1] *Mohib Ullah*[2] *Maqsood Mahmud*[3] *Faouzi Alaya Cheikh*[2]

[1] College of Computer Science and Engineering, University of Ha'il, Saudi Arabia.
[2] Department of Computer Science, Norwegian University of Science and Technology, Norway.
[3] Department of Management Information System,
University of Imam Abdulrahman Bin Faisal, Dammam, Saudi Arabia.

ABSTRACT

Generative adversarial networks (GANs) have been significantly investigated in the past few years due to its outstanding data generation capacity. The extensive use of the GANs techniques is dominant in the field of computer vision, for example, plausible image generation, image to image translation, facial attribute manipulation, improving image resolution, and image to text translation. In spite of the significant success achieved in these domains, applying GANs to various other problems still presents important challenges. Several reviews and surveys for GANs are available in the literature. However, none of them present short but focused review about the most significant aspects of GANs. In this paper, we address these aspects. We analyze the basic theory of GANs and the differences among various generative models. Then, we discuss the recent spectrum of applications covered by the GANs. We also provide an insight into the challenges and future directions.

Index Terms— Generative adversarial networks, Laplacian pyramid of adversarial networks, deep convolutional GAN, boundary equilibrium GAN, progressive GAN.

1. INTRODUCTION

The field of machine learning has grown significantly in the past decade. Fields ranging from finance [1] to security [2], from health-care [3] to people safety [4], from marketing [5] to autonomous vehicles [6], all make use of machine learning techniques. Therefore, huge amount of work in both academia and industry is carried out to develop new machine learning techniques. Research papers related to speech recognition, video surveillance, natural language processing, object recognition, and remote sensing are published almost every day. Specifically, in the field of computer vision and image classification, the contributions of research are much higher in recent years.

Two popular machine learning techniques are generative models and discriminative models. It is worth noticing that the usage of generative models as compared to discriminative models was not significant due to the difficulty of estimating various probabilistic parameters. However, Goodfellow et al. [7] overcame these key challenges by investigating generative adversarial networks (GANs), which are an emerging technique for both semi-supervised and unsupervised learning. A GAN model achieves this through modeling high-dimensional distributions of data characterized by training a pair of networks in competition with each other.

GANs computes a density function over a data distribution considering a set of different training approaches. The main concept of a GAN is to train two networks: a generator (G) and a discriminator (D), in a competition of minimax game. The purpose of the generator is to produce realistic images that can cheat the discriminator. The discriminator attempts to classify engendered images (produced by the generator part of the network) as forgery and classifies the real images from the original samples as real. The pair of networks in competition with each other in the form of the minimax game is an auxiliary and congenital means of computing the density function of the original sample images. The GANs models can produce samples in parallel without exploiting run-time proportional to the dimensionality of input. Comparing to Boltzmann machines, the modeling of the generator function presents limited restrictions. A few probability distributions admit tractable Markov chain sampling in case of Boltzmann machines. The GANs models do not require Markov chains and variational bound. Despite these advantages, the GANs models need to compute the Nash equilibrium of a game for the purpose of training. It is important to note that it is significantly complicated than optimizing an objective function.

In general, deep learning techniques consider some form of stochastic approximation. These sampling-based approximations perform better if a fair sample can be produced swiftly. Other models need to produce expensive samples by exploiting algorithms based on Markov chain. However, the convergence of Markov chain can be very slow and there is no consolidated approach to find out the convergence of the chain. Boltzmann machines are generative models based on Markov chains. Nowadays, Boltzmann machines are not used frequently since Markov chain approximation techniques are not scaled to challenges like ImageNet generation. Therefore, GANs are investigated to ignore Markov chains for these reasons.

GANs have been used in many application driven tasks including video frame prediction [8], abnormal event detection [9], improving image resolution [10], generative image manipulation [11], visual tracking [12], and image to text translation [13]. In all these applications and many others not mentioned, GANs have shown significant progress. GANs have successfully shown to be the state-of-the-art for producing sharp and realistic images for many applications.

The rest of the paper presents GANs and variants in Section 2, various applications areas in Section 3, challenges and future direction in Section 4. In Section 5, the conclusion of the paper is presented.

2. GANS AND VARIANTS

Some generative models are based on the principle of maximum likelihood [14]. The generative models that do not use maximum likelihood by default can be made to do so. We consider the maximum likelihood version of generative models that do not generally use maximum likelihood. Therefore, we can ignore some of the more distracting differences between different generative models.

The fundamental concept of maximum likelihood is to develop a model that provides an estimate of a probability distribution, parameterized by parameters θ. We then consider the likelihood as the probability that the model assigns to the training data, for a dataset consisting of m training samples. The principle of maximum likelihood identifies the parameters for the model that maximize the likelihood of the training data. In term of computation, it is better to do it in log space, where we perform computations in the form of a sum rather than a product over examples. This way of formulations simplify the expressions for the derivatives of the likelihood with respect to the models. One can think of maximum likelihood estimation as minimizing the KullbackLeibler (KL) divergence between the data generating distribution and the model.

There are many different types of GANs architecture-variants introduced in the literature which are mainly proposed for the purpose of different applications. The original GAN paper [7] considers fully-connected neural networks for both parts namely generator and discriminator. Therefore, it is fully-connected GAN (FCGAN). This GAN is evaluated using simple datasets including MNIST [15], CIFAR-10 [16] and Toronto Face Dataset. It does not present good generalization performance for very complex types of images. Denton et al. [17] proposed Laplacian pyramid of adversarial networks (LAPGAN) capable of producing high quality samples of natural images. Their method uses a cascade of convolutional networks (convnets) within a Laplacian pyramid framework to generate images in a coarse-to-fine fashion. At each level of the pyramid a separate generative convnet model is trained using the generative adversarial nets technique. Samples drawn from their model are of significantly higher quality than existing models. Radford et al. [18] proposed deep convolutional GAN (DCGAN) to fill the gap between the success of CNNs for supervised learning and unsupervised learning. Their networks (DCGANs) have certain architectural constraints. They showed that DCGAN is a strong candidate for unsupervised learning. Training on various image datasets, they showed that their model pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. They also exploit the learned features for novel tasks - demonstrating their applicability as general image representations. Zhao et al. [19] introduced boundary equilibrium GAN (BEGAN) that uses an autoencoder architecture for the discriminator. In competition to traditional optimization, the BEGAN matches the autoencoder loss distributions using a loss derived from the Wasserstein distance instead of matching data distributions directly. This contribution helps [18] to produce easy to- reconstruct data for the autoencoder at the beginning because the generated data is close to 0 and the real data distribution has not been learned accurately yet, which prevents D easily winning [18] at the early training stage. Karras et al. [20] introduced progressive GAN (PROGAN). The concept is to grow both the generator and discriminator progressively. The model starts from a low resolution and adds new layers that model increasingly fine details as training progresses. This model speeds the training up, greatly stabilizes it, and produces images of unprecedented quality. Zhang et al. [21] proposed the self-attention generative adversarial network (SAGAN) which allows attention-driven, long-range dependency modeling for image generation problems. Traditional convolutional GANs produce high-resolution details as a function of only spatially local points in lower-resolution feature maps. In SAGAN, details can be produced considering cues from all feature locations. Furthermore, the discriminator can check that highly informative features in distant portions of the image are consistent with each other. Brock et al. [22] proposed big GAN (BigGANs) that achieved state-of-the-art performance on the ImageNet datasets. They found that applying orthogonal regularization to the generator renders it amenable to a simple ”truncation trick,” allowing fine control over the trade-off between sample fidelity and variety by reducing the variance of the Generator’s input. Their contributions lead to models which set the new state of the art in class-conditional image synthesis.

3. APPLICATIONS

GANs have been used in many application areas including image-to-image translation, text-to-image translation, semantic-image-to-photo translation, face frontal view generation, photos to emojis, and super resolution. However, in this section, we will focus on areas where GANs are adopted recently. These include abnormal event detection, action recognition, pose estimation, and depth estimation.

3.1. Abnormal Event Detection

Ravanbakhsh et al. [23] proposed generative adversarial nets trained on normal frames and corresponding optical-flow images in order to learn an internal representation of a scene depicting normal situation. The proposed GANs are not able to generate abnormal events since they are trained considering only normal data. During the testing stage, the real data are compared with both the appearance and the motion representations reconstructed by the GANs and abnormal areas are detected by computing local differences. Lee et al. [24] introduced a spatio-temporal generator which synthesizes an inter-frame by using spatio-temporal characteristics with bidirectional ConvLSTM. The spatio-temporal discriminator finds whether an input sequence is real-normal or not with 3D convolutional layers. They trained the two networks in an adversarial way to effectively formulate spatio-temporal features of normal patterns. After the learning stage, the generator and the discriminator can be independently used as detectors, and deviations from the learned normal patterns are detected as abnormalities. Wang et al. [25] investigated that deep generative models have the risk of overfitting training samples, which has negative effects on anomaly detection. To handle this key challenge, they propose a self-adversarial variational autoencoder with a Gaussian anomaly prior assumption. They assumed that both the anomalous and the normal prior distributions are Gaussian and have overlaps in the latent space. Therefore, they trained a Gaussian network T to synthesize anomalous and near-normal latent variables. Keeping the original training objective of variational autoencoder, besides, the generator G tries to distinguish between the normal latent variables and the anomalous ones synthesized by T, and the encoder E is trained to discriminate whether the output of G is real. Khan et al. [26] detected falls by training the classifier on only the normal activities and identifying a fall as an anomaly. For this purpose, they use an adversarial learning framework consisting of a spatio-temporal autoencoder for reconstructing input video frames and a spatio-temporal convolution network to discriminate them against original video frames. They use 3D convolutions to learn spatial and temporal features from the input video frames. The adversarial learning of the spatio-temporal autoencoder effectively reconstructs the normal activities and detects unseen falls plausible within this framework. Lei et al. [27] combined spatial and temporal features to model the input normal pattern. They fuse two AutoEncoders: one is trained to learn spatial features and the other is trained to learn temporal features. The output of the spatial network and input is fed to the temporal network. Anomaly is detected by taking the difference between a prediction of future frame and its ground truth. Li et al. [28] introduce an unsupervised multivariate anomaly detection method based on generative adversarial networks (GANs), using the long-short-term-memory recurrent neural networks (LSTM-RNN) as the base models (namely, the generator and discriminator) in the GAN framework to encode the temporal correlation of time series distributions. Their method uses the entire variable set concurrently to encode the latent interactions amongst the variables.

3.2. Action Recognition

Wang et al. [29] introduced an end-to-end architecture that improves the discriminability of features of partially observed videos by assimilating them to features from complete videos. They use the generative adversarial network, which improves the recognition accuracy of partially observed videos through narrowing the feature difference of partially observed videos from complete ones. Their generator consists of two networks: a CNN for feature extraction and an LSTM for estimating residual error between features of the partially observed videos and complete ones, and then the features from CNN adds the residual error from LSTM, which is regarded as the improved feature to fool a competing discriminator. Gammulle et al. [30] proposed a conditional GAN (cGAN) model for continuous fine-grained human action segmentation using multi-modal data and learned scene context information. Their method consists of two GANs: termed action GAN and auxiliary GAN. The action GAN is trained to operate over the current RGB frame and the auxiliary GAN considers supplementary information such as depth or optical flow. The goal of both GANs is to produce similar 'action codes', a vector representation of the current action. Shou et al. [31] proposed a generator network, which reduces noises in motion vectors and encodes fine motion details, achieving a more discriminative motion cue (DMC) representation for action recognition. Alnujaim et al. [32] produced a large number of micro-Doppler signatures using GANs to increase the training data to classify human activities. For each human activity, corresponding GANs that produce micro-Doppler signatures for a particular activity are constructed. Dwivedi et al. [33] investigated ProtoGAN framework which synthesizes additional action examples for novel categories by conditioning a conditional generative adversarial network with class prototype vectors. These vectors are learnt considering a class prototype transfer network (CPTN) from examples of seen categories. Their synthesized examples for a novel class are semantically similar to real examples belonging to that class and is used to train a model exhibiting better generalization towards novel classes.

3.3. Pose Estimation

Wang et al. [34] applied self-attention GAN to improve the performance of human pose estimation. With attention mechanism in the framework of GAN, they can learn long-range body joints dependencies, therefore, enforce the entire body joints structural constrains to make all the body joints to be consistent. Zhu et al. [35] proposed hard joints mining method, for human pose estimation, based on the generative adversarial network, which consists of two stacked hourglasses with a similar architecture. During the training period, the discriminator distinguishes the generated heatmaps from the ground-truth heatmaps and introduces the adversarial loss to the generator through back-propagation to induce generator produces a more reasonable prediction. For the same problem, Peng et al. [36] designed a generator that is an augmentation network competing against a discriminator (e.g. a target network) by producing hard examples online. The generator exploits weaknesses of the discriminator, and the discriminator learns from hard augmentations to achieve better performance. Yang et al. [37] modeled an adversarial learning framework, which distills the 3D human pose structures learned from the fully annotated dataset to in-the-wild images with only 2D pose annotations. They designed a novel multi-source discriminator to distinguish the predicted 3D poses from the ground truth, which helps to enforce the pose estimator to produce anthropometrically valid poses even with images in the wild.

3.4. Depth Estimation

Feng et al. [38] introduced stacked generative adversarial network (SGANVO) for visual depth and ego-motion estimation. The network consists of a stack of GAN layers, where lowest layers estimate the depth and ego-motion and the higher layers estimate the spatial features. Arslan and Seke [39] developed a GAN-based method for depth map estimation from any given single face image. Many variants of GANs have been tested for the depth estimation. They evaluated that conditional Wasserstein GAN structure offers the most robust approach. Ji et al. [40] addressed the problem of monocular depth estimation when only a limited amount of training image-depth pairs are available. Aiming to break the bottleneck of expensive data collections, they investigated a semi-supervised adversarial learning framework, which only utilizes a small amount of image-depth pairs with a large amount of cheaply-available monocular images to pursuit high accuracy. They use one generator to regress the depth and two discriminators to evaluate the predicted depth. These two discriminators provide their feedbacks to the generator as the loss to produce more realistic and accurate depth predictions. Aleotti et al. [41] proposed to cast unsupervised monocular depth estimation within a GAN paradigm. The generator network learns to infer depth from the reference image to produce a warped target image. At training time, the discriminator network learns to distinguish between fake images generated by the generator and target frames acquired with a stereo rig. Kumar et al. [42] presented a technique for monocular reconstruction, i.e. depth map and pose prediction from input monocular video sequences, using adversarial learning. They extended geometry-aware neural network architectures that learn from photoconsistency-based reconstruction loss functions defined over spatially and temporally adjacent images by leveraging recent advances in adversarial learning. They introduced a GAN that can learn improved reconstruction models, with flexible loss functions that are less susceptible to adversarial examples, using generic semi-supervised or unsupervised datasets. The generator function in the investigated GAN learns to synthesize neighboring images to predict a depth map and relative object pose, and the discriminator function learns the distribution of monocular images to correctly classify the authenticity of the synthesized images.

4. CHALLENGES AND FUTURE DIRECTIONS

As we have discussed different application areas, GANs go beyond recognition and classification and, as the name implies, actually produce output based on a reference or sample. In fact, the outcomes of GANs are convincing variations of reality based on real images or streams of frames in a video sequence. However, in term of functionality, any GAN is not much different than other convolutional neural networks. The main formulation of the discriminator in a GAN is similar to a basic image classifier and the generator is also similar to other convolutional neural networks.

Considering GANs, there is the key challenge of mode collapse, which endangers the stability of that training and feedback stages. Essentially, one network can subdue the other. For example, the generator might produce images the discriminator cannot assess, even when those engendered images might not look like they should based on the dataset distribution. Therefore, the generator will never learn because it is not getting feedback about what to do better. The GAN overpowering problem can be overcomed over time. Both software and hardware challenges need to be addressed before GANs at any level can get into areas beyond image and video generation and into broader applications considering scientific, technical, or enterprise realms. The GANs results are compelling in image and video but they are not very compelling in text and audio applications. However, researchers might figure things out in the near future. So the most success is in visual domain and the same is happening in medical imaging.

Training a GAN is training two networks, generator and discriminator. They both try to race against each other and try to reach an optimum point. In fact, the generator and discriminator reach a state where they cannot improve further. Therefore, the configuration of gradient descent tries to reduce the loss function defined on the problem. However, the setup is by no means enforcing the networks to reach optimum point, which have non-convex objective with continuous high dimensional parameters. Both the networks try to take steps one after another to minimize a non-convex objective, however, the process ends up in an oscillation rather than reducing the underlying true objective. In practice, when the discriminator achieves a loss very close to zero, then there is something wrong with the model and figuring it out is a challenging issue.

The use of GANs to explore and handle other application areas is yet another future direction. These application areas include but not limited to crowd motion analysis, remote sensing, and hyper-spectral image analysis. Crowd motion analysis is prone to several challenging tasks including motion segmentation [43], crowd counting [44], multiple target tracking [45], congestion detection and localization, rare behavior simulation, and stampede detection to name a few. To cope with these problems, GANs present better platform. For example, acquiring labeled data for rare crowd motion patterns is a complicated process. Therefore, GANs can be best explored since they don't require labeled data. They can be trained using unlabeled data as they learn the internal representations of the data. Another important strength of GANs that could be effective to use for crowd motion analysis is that they generate data that is similar to real data. Therefore, they could be modeled for crowd simulation applications. Since GANs learn the internal representations of data, they could be adopted to detect congestion considering messy and complicated distributions of crowd data. After training, the discriminator network is a classifier and can be used to classify crowd behaviors [46].

5. CONCLUSION

This paper presents the background of GANs, discusses the architecture and variants, and explores new application areas. In addition, the challenges of GANs and future directions are presented.

6. REFERENCES

[1] Kang Zhang, Guoqiang Zhong, Junyu Dong, Shengke Wang, and Yong Wang, "Stock market prediction based on generative adversarial network," *Procedia computer science*, vol. 147, pp. 400–406, 2019.

[2] Muhammad Mudassar Yamin, Basel Katt, and Vasileios Gkioulos, "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture," *Computers & Security*, p. 101636, 2019.

[3] Ahmed Kedir, Mohib Ullah, and Bauer Jacob R, "Spectranet: A deep model for skin oxygenation measurement from multi-spectral data," *Electronic Imaging Conference*, 2020.

[4] Habib Ullah, Ahmed B Altamimi, Muhammad Uzair, and Mohib Ullah, "Anomalous entities detection and localization in pedestrian flows," *Neurocomputing*, vol. 290, pp. 74–86, 2018.

[5] Germán Sánchez-Hernández, Francisco Chiclana, Núria Agell, and Juan Carlos Aguado, "Ranking and selection of unsupervised learning marketing segmentation," *Knowledge-based systems*, vol. 44, pp. 20–33, 2013.

[6] Stefan Hoermann, Martin Bach, and Klaus Dietmayer, "Dynamic occupancy grid prediction for urban autonomous driving: A deep learning approach with fully automatic labeling," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2056–2063.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[8] William Lotter, Gabriel Kreiman, and David Cox, "Deep predictive coding networks for video prediction and unsupervised learning," *arXiv preprint arXiv:1605.08104*, 2016.

[9] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regazzoni, and Nicu Sebe, "Abnormal event detection in videos using generative adversarial nets," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 1577–1581.

[10] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[11] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A Efros, "Generative visual manipulation on the natural image manifold," in *European Conference on Computer Vision*. Springer, 2016, pp. 597–613.

[12] Yibing Song, Chao Ma, Xiaohe Wu, Lijun Gong, Linchao Bao, Wangmeng Zuo, Chunhua Shen, Rynson WH Lau, and Ming-Hsuan Yang, "Vital: Visual tracking via adversarial learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8990–8999.

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[14] Ian Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *arXiv preprint arXiv:1701.00160*, 2016.

[15] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] Alex Krizhevsky, Geoffrey Hinton, et al., "Learning multiple layers of features from tiny images," Tech. Rep., Citeseer, 2009.

[17] Emily L Denton, Soumith Chintala, Rob Fergus, et al., "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in neural information processing systems*, 2015, pp. 1486–1494.

[18] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[19] Junbo Zhao, Michael Mathieu, and Yann LeCun, "Energy-based generative adversarial network," *arXiv preprint arXiv:1609.03126*, 2016.

[20] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[21] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena, "Self-attention generative adversarial networks," *arXiv preprint arXiv:1805.08318*, 2018.

[22] Andrew Brock, Jeff Donahue, and Karen Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.

[23] Mahdyar Ravanbakhsh, Enver Sangineto, Moin Nabi, and Nicu Sebe, "Training adversarial discriminators for cross-channel abnormal event detection in crowds," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1896–1904.

[24] Sangmin Lee, Hak Gu Kim, and Yong Man Ro, "Stan: spatio-temporal adversarial networks for abnormal event detection," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1323–1327.

[25] Xuhong Wang, Ying Du, Shijie Lin, Ping Cui, and Yupu Yang, "Self-adversarial variational autoencoder with gaussian anomaly prior distribution for anomaly detection," *arXiv preprint arXiv:1903.00904*, 2019.

[26] Shehroz S Khan, Jacob Nogas, and Alex Mihailidis, "Spatio-temporal adversarial learning for detecting unseen falls," *arXiv preprint arXiv:1905.07817*, 2019.

[27] Zheng Lei, Fang Deng, and Xudong Yang, "Spatial temporal balanced generative adversarial autoencoder for anomaly detection," in *Proceedings of the 2019 International Conference on Image, Video and Signal Processing*. ACM, 2019, pp. 1–7.

[28] Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng, "Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks," in *International Conference on Artificial Neural Networks*. Springer, 2019, pp. 703–716.

[29] Dong Wang, Yuan Yuan, and Qi Wang, "Early action prediction with generative adversarial networks," *IEEE Access*, vol. 7, pp. 35795–35804, 2019.

[30] Harshala Gammulle, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes, "Coupled generative adversarial network for continuous fine-grained action segmentation," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 200–209.

[31] Zheng Shou, Xudong Lin, Yannis Kalantidis, Laura Sevilla-Lara, Marcus Rohrbach, Shih-Fu Chang, and Zhicheng Yan, "Dmc-net: Generating discriminative motion cues for fast compressed video action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1268–1277.

[32] Ibrahim Alnujaim, Daegun Oh, and Youngwook Kim, "Generative adversarial networks for classification of micro-doppler signatures of human activity," *IEEE Geoscience and Remote Sensing Letters*, 2019.

[33] Sai Kumar Dwivedi, Vikram Gupta, Rahul Mitra, Shuaib Ahmed, and Arjun Jain, "Protogan: Towards few shot learning for action recognition," *arXiv preprint arXiv:1909.07945*, 2019.

[34] Xiangyang Wang, Zhongzheng Cao, Rui Wang, Zhi Liu, and Xiaoqiang Zhu, "Improving human pose estimation with self-attention generative adversarial networks," *IEEE Access*, vol. 7, pp. 119668–119680, 2019.

[35] Aichun Zhu, Sai Zhang, Yaoying Huang, Fangqiang Hu, Ran Cui, and Gang Hua, "Exploring hard joints mining via hourglass-based generative adversarial network for human pose estimation," *AIP Advances*, vol. 9, no. 3, pp. 035321, 2019.

[36] Xi Peng, Zhiqiang Tang, Fei Yang, Rogerio S Feris, and Dimitris Metaxas, "Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2226–2234.

[37] Wei Yang, Wanli Ouyang, Xiaolong Wang, Jimmy Ren, Hongsheng Li, and Xiaogang Wang, "3d human pose estimation in the wild by adversarial learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5255–5264.

[38] Tuo Feng and Dongbing Gu, "Sganvo: Unsupervised deep visual odometry and depth estimation with stacked generative adversarial networks," *arXiv preprint arXiv:1906.08889*, 2019.

[39] Abdullah Taha Arslan and Erol Seke, "Face depth estimation with conditional generative adversarial networks," *IEEE Access*, vol. 7, pp. 23222–23231, 2019.

[40] Rongrong Ji, Ke Li, Yan Wang, Xiaoshuai Sun, Feng Guo, Xiaowei Guo, Yongjian Wu, Feiyue Huang, and Jiebo Luo, "Semi-supervised adversarial monocular depth estimation," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[41] Filippo Aleotti, Fabio Tosi, Matteo Poggi, and Stefano Mattoccia, "Generative adversarial networks for unsupervised monocular depth prediction," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[42] Arun CS Kumar, Suchendra M Bhandarkar, and Mukta Prasad, "Monocular depth prediction using generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 300–308.

[43] Habib Ullah and Nicola Conci, "Crowd motion segmentation and anomaly detection via multi-label optimization," in *ICPR workshop on pattern recognition and crowd analysis*, 2012.

[44] Saleh Basalamah, Sultan Daud Khan, and Habib Ullah, "Scale driven convolutional neural network model for people counting and localization in crowd scenes," *IEEE Access*, 2019.

[45] Mohib Ullah, Habib Ullah, and Faouzi Alaya Cheikh, "Single shot appearance model (ssam) for multi-target tracking," *Electronic Imaging*, vol. 2019, no. 7, pp. 466–1, 2019.

[46] Mohib Ullah, Habib Ullah, Nicola Conci, and Francesco GB De Natale, "Crowd behavior identification," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 1195–1199.